Before the

Federal Communications Commission

Washington, DC 20554

In the Matter of	)	
Preserving the Open Internet	)	GN Docket No. 09-191
Broadband Industry Practices	)	WC Docket No. 07-52

Comments of Christopher S. Yoo

Professor of Law, Communication, and Computer & Information Science
Founding Director of the Center for Technology, Innovation and Competition
University of Pennsylvania

INTRODUCTION

I offer my views as to the manner in which the proposed Open Internet rules should take specialized services and mobile wireless Internet access services into account. I also offer four principles to guide the application of a case-by-case approach to enforcing the proposed Open Internet regulations. The views presented are my own and should not be attributed to the University of Pennsylvania or the Center for Technology, Innovation and Competition.[1]

I. SPECIALIZED SERVICES

Network providers have employed specialized services in a myriad of ways that benefit consumers. The ability to reserve bandwidth and prioritized traffic has allowed them to use

[1] Broadband for America has commissioned me to offer a legal opinion as to whether reclassifying broadband Internet access as a Title II telecommunications service would survive judicial scrutiny. I have not been engaged by any party to address the two underdeveloped issues on which the Commission has sought comment in its Public Notice of September 1, 2010.

bandwidth more efficiently, offer better service, and reduce costs. The cost reductions made possible by specialized services reduces the number of customers a system needs to breakeven, which is particularly important in view of the difficulty in building out high-speed network capability identified by the National Broadband Plan.

The following section reviews seven examples of specialized services that have created real benefits for consumers. These are precisely the type of practices that the Commission must be careful not to foreclose as it conducts its Open Internet proceeding.

A. Prioritization of Terminal Sessions over File Transfer on the NSFNET

One of the earliest examples of a specialized service occurred on the NSFNET in 1987, when end users first began to connect to the network through personal computers (PCs) instead of dumb terminals.[2] Terminal sessions are an extremely interactive application, in which every key stroke is immediately transmitted and which requires constant, real-time interaction with the network. Any delay causes the terminal to lock up temporarily. File transfers are much less interactive. Particularly given the 56 kbps backbone speeds of the time, end users would typically expect such applications to last several minutes.

The advent of PCs made it much easier for end users to transfer files, which in turn increased the intensity of the demands that end users were placing on the network to the point where the network slowed to a crawl. The resulting congestion caused terminal sessions to run agonizingly slowly, and the fact that fixed cost investments cannot be made instantaneously created an unavoidable delay in adding network capacity.

---

[2] *See* Christopher S. Yoo, *Beyond Network Neutrality*, 19 HARV. J.L. & TECH. 1, 22-23 (2005). For other descriptions, see Robert M. McDowell, *Who Should Solve This Internet Crisis*, WASH. POST, July 28, 2008, at A17; Roger Bohn et al., *Mitigating the Coming Internet Crunch: Multiple Service Levels via Precedence*, _ J. HIGH SPEED NETWORKING _ (1993); Jeffery K. MacKie-Mason & Hal R. Varian, *Economic FAQs About the Internet*, 1 J. ELEC. PUBL'G (1995), http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;idno=3336451.0002.102;cc=jep;rgn=main;view=text.

NSFNET’s interim solution was to reprogram its routers to give traffic running the application protocol associated with terminal sessions (telnet) higher priority than traffic running the application associated with file transfer sessions (File Transfer Protocol or FTP) until additional bandwidth could be added. In short, intelligence in the core of the network looked inside packets and gave a higher priority to interactive, real-time traffic and deprioritized traffic that was less sensitive to delay. The network also made wider use of prioritization in the type of service field in the IP header.

This episode demonstrates why forecasting the amount of network capacity is so difficult. The spike in traffic was driven not by any change within the network itself, but rather a major innovation in a complementary technology (the PC) that changed the ways people used the network. In this sense, it bears a striking resemblance to the state of affairs in 1995 and 1996, when the simultaneous development of HTML and Mosaic, the first graphically oriented browser, caused Internet traffic to accelerate to a rate of 800% to 900% and to turn the network into what many dubbed “the World Wide Wait.” As difficult as it is to correctly anticipate developments within the network, it is even harder to foresee game-changing improvements in complementary technologies.

This episode also demonstrates the beneficial role that network management can play in providing a better end user experience. Indeed, prioritization actually might have been able to offer better service to users of terminal sessions without degrading the experience of file transfer users. This is because the performance of file transfer sessions depends entirely on when the last packet arrives. Interactive applications (particularly streaming applications), in contrast, are very sensitive to the speed and spacing with which intermediate packets arrive. So long as the delivery time of the last packet is not affected, the network can rearrange the delivery schedule for intermediate packets associated with terminal sessions without adversely affecting overall performance file transfer sessions.

At the same time, this episode demonstrates how core-based solutions that explicitly route traffic based on the application layer protocol with which it is associated can benefit consumers. Although this example represented a short-run solution, in theory such solutions need not be temporary. Indeed, in a technologically dynamic world, one would expect at times that employing network management techniques would be cheaper than adding bandwidth, and vice versa. Moreover, one would also expect the relative cost of these alternative solutions (and the balance that they imply) to change over time.

B. Public Safety Spectrum in 700 MHz Auction

Another prime example of the benefits of specialized services appeared in the rules that the FCC established to govern the public safety spectrum contained in the D Block of the 700 MHz auction. The rules required that the licensee build a single network shared both by public safety users and commercial users. Public safety users would be given the unconditional right to preempt commercial traffic during emergency situations. Commercial users would operate on a secondary basis that must accept interference form primary users at all times and must not interfere with the primary users. In essence, the rules established two classes of service operating on the same network, with the primary, higher value use being given unconditional priority over all secondary uses.[3]

The Commission based its decision on three rationales. First, the higher utilization form combining two different types of uses can reduce costs by allowing the network to realize

---

[3] *Service Rules for the 698-746, 747-762 and 777-792 MHz Bands*, Second Report and Order, 22 FCC Rcd. 15289, 15437 ¶ 414 (2007).

economies of scale. Second, sharing spectrum with multiple users promotes more efficient use of public safety spectrum. Third, the addition of secondary uses will help defray the cost of building out the network without adversely affecting the needs of the higher value, primary uses.[4]

The FCC's decisions with respect to the D Block in the 700 MHz auction provide a powerful exposition of the benefits of allowing networks to give priority to higher value traffic traveling in the same pipe as lower value traffic. Doing so both yields consumer benefits and promotes competition in areas where broadband is already available. Making it easier for network providers to cover the costs of constructing new networks also promotes and entry into currently unserved areas.

C. Peha's Real-Time Secondary Markets for Spectrum

Another interesting example of a specialized service was proposed by Jon Peha in a paper co-authored with one of his graduate students before he became the FCC's Chief Technologist. The paper takes as its starting point the classic tradeoff between licensed and unlicensed uses of spectrum. Because exclusive licensing typically restricts access to a limited number of users, it tends to use spectrum inefficiently when those users connect to the network sporadically and the resource lays fallow whenever those particular parties are not using the network. At the same time, exclusive licensing does enable the network to offer guaranteed levels of quality of service. Unlicensed spectrum reverses this tradeoff. The fact that any number of users can share the same spectrum allows the resource to be used more efficiently. At the same time, unlicensed spectrum is unable to provide guaranteed levels of quality of service, in part because the

[4] *Id.* at 15431 ¶ 396, 15437 ¶ 416; *Implementing a Nationwide, Broadband, Interoperable Public Safety Network in the 700 MHz Band*, Ninth Notice of Proposed Rulemaking, 21 FCC Rcd. 14837, 14847 ¶ 40, 14848 ¶ 42 (2006).

openness of the resource provides little incentive to conserve the amount of bandwidth used and in part because there is no way to limit the number of devices connected to the network.[5]

Peha proposed a hybrid system that "offer[s] both the efficiencies of sharing with the possibility of quality of service guarantees."[6] Under this approach, spectrum is exclusively licensed. Secondary users can request permission to share the spectrum for a fee, but the license holder would be allowed to deny access if the network is already saturated with prior calls.[7] The ability to generate additional revenue without degrading existing sources of revenue makes it easier for networks to breakeven, which should promote buildout and help alleviate the digital divide. Although secondary users would receive lower priority, they benefit from paying prices estimated to be only one third of those paid by primary users.[8]

In essence, this proposal is simply a form of prioritized service, in which the provider divides a single pipe into two tiers, each offering different levels of quality of service and different prices. The addition of a lower quality, lower price tier allows for more efficient use of the spectrum and makes service more affordable by allowing it to be offered at a lower price point. At the same time, it provides consumers who wish to run more demanding applications with the choice of a service able to offer better quality of service guarantees.

D. PlusNet

The innovative network management techniques employed by British DSL provider PlusNet provide another example of the potential benefits of specialized services.[9] PlusNet uses

---

5 Jon M. Peha & Sookan Panichpapiboon, *Real-Time Secondary Markets for Spectrum*, 28 TELECOMM. POL'Y 603, 604 (2004).

6 *Id.*

7 *Id.* at 604-05.

8 *Id.* at 612.

9 *See* Carol Wilson, *The Consumer Friendly Version of DPI*, TELEPHONYONLINE, July 29, 2008, *currently available at* http://connectedplanetonline.com/home/news/dpi-consumer-friendly-0729/; *Transparency the Key to*

deep packet inspection (DPI) to divide the data stream into multiple different levels of priority.[10] In so doing, PlusNet has served as a model of public disclosure, explaining how what it is doing to prioritize traffic, why connection speeds vary in particular cases,[11] and offering meaningful guidance as to expected speeds during different times of day.[12] Prioritizing traffic in this manner has enabled PlusNet to win numerous industry awards for the quality of their network connections and for customer satisfaction.[13]

In many ways, DPI has generated undeserved criticism. Sometimes denigrated as a deviation from network norms, DPI is widely used by most (if not all) major ISPs to examine of samples of traffic to search for security threats. PlusNet provides a particularly telling example of why a reflexive hostility toward DPI is unwarranted.

E. AT&T's U-verse

As the Commission noted when first seeking comment on its Open Internet proposal in October 2009, AT&T's U-verse represents an important example of a specialized service.[14] The benefits provided by prioritization and reserving bandwidth are vividly illustrated by comparing U-Verse with Verizon's FiOS network. Verizon is investing $23 billion to build its fiber-based FiOS network, which offers up to 100 Mbps and holds the promise of providing up to 10 Gbps of service. Instead of creating a new last-mile network, AT&T's strategy leverages the existing

---

PlusNet's Success, STAR (Sheffield, UK), Apr. 9, 2008, http://www.thestar.co.uk/business/Transparency-the-key-to-PlusNet39s.3961139.jp.

[10] *Broadband: Traffic Prioritisation*, PLUSNET, Apr. 15, 2010, http://www.plus.net/support/broadband/quality_broadband/traffic_prioritisation.shtml.

[11] *Broadband: Your Broadband Speed - The Basics*, PLUSNET, July 19, 2010, http://www.plus.net/support/broadband/speed_guide/speed_basics.shtml.

[12] *Broadband: Broadband Download Speeds*, PLUSNET, May 25, 2010, http://www.plus.net/support/broadband/speed_guide/download_speeds.shtml.

[13] *Award Winning Broadband and Quality Customer Service*, PLUSNET, http://www.plus.net/press/awards.shtml.

[14] *Preserving the Open Internet*, Notice of Proposed Rulemaking, 24 FCC Rcd. 13064, 13116-17 ¶ 150 (2009).

infrastructure by deploying a VDSL2-based technology. U-verse provides smaller amounts of bandwidth, ranging from 20 to 32 Mbps depending on a particular customer's location, but at the much lower cost of $6 to $7 billion.

The problem is that U-verse does not have enough bandwidth to provide video in the same manner as cable companies and FiOS. Thus instead of sending all of the available channels all the time regardless of whether anyone is watching, U-verse uses a switched digital video technology that only transmits the particular channel that the subscriber is watching at any particular time. In addition, it avoids the delays that can render video programming unwatchable by giving a higher priority to the traffic associated with its own proprietary video offerings.

In many ways, AT&T's practices represent precisely the type of conduct that gives network neutrality proponents pause. It prioritizes a single application (video) from a single source (AT&T) and runs the risk of allowing AT&T to gain a competitive advantage by favoring its own content over others. And yet, these practices are what has allowed AT&T to avoid having to spend an additional $17 billion needed to deploy fiber-based solutions like FiOS.

Given the magnitude of the looming problem and the tightening of the capital markets associated with the ongoing recession, policymakers should avoid regulations that make higher capital investments the only solution to the problem of video-induced traffic growth and should instead permit networks to use prioritization to employ more efficiently the capacity that already exists. Placing regulatory restrictions on network management would not only degrade the service of existing customers. Increasing the amount of capacity needed to support a particular number of customers would increase the per capita expense of building new networks. This de facto increase in costs would both limit broadband deployment in rural and other low-density populations, as demonstrated by the large number of filings by public officials and business leaders from rural areas and small towns opposing the Open Internet initiative.

F. Internet2's Interoperable On-demand Network (ION)

One of the central tenets underlying the Internet is that routers should operate on a pure store-and-forward basis without having to keep track of what happens to packets after they have been passed on. This commitment is reflected in the Internet's general hostility toward virtual circuits and the belief that routers should not maintain per-flow state. Opponents of network management often point to the Senate testimony offered by officials of Internet2—a nonprofit partnership of universities, corporations, and other organizations devoted to advancing the state of the Internet—noting that, although their network designers initially assumed that ensuring quality of service required building intelligence into the network, "all of [their] research and practical experience supported the conclusion that it was far more cost effective to simply provide more bandwidth."[15]

To a certain extent, this longstanding hostility toward virtual circuits is an artifact of the Internet's military origins that has less relevance for the Internet of today. DARPA protocol architect David Clark has pointed out that the belief that routers operating in the core of the network should not maintain per-flow state derived largely from the high priority that military planners placed on survivability.[16] Clark notes, however, that survivability does not represent a significant concern for the modern Internet. Moreover, technologies such as IntServ[17] and

---

15 *Net Neutrality: Hearing Before the S. Comm. on Commerce, Science, and Transportation*, 109th Cong. 64, 66 (2006) (statement of Gary R. Bachula, Vice President, External Affairs, Internet2), *available at* http://commerce.senate.gov/public/?a=Files.Serve&File_id=c5bf9e54-b51f-4162-ab92-d8a6958a33f8.

16 Clark, *supra* note 2, at 107-08.

17 Robert Braden et al., *Integrated Services in the Internet Architecture: An Overview*, IETF RFC 1633 (rel. July 1994), *available at* http://www.rfc-editor.org/rfc/rfc1633.pdf.

MPLS,[18] both of which are governed by accepted IETF standards, employ what amount to virtual circuits to enhance quality of service and to increase network efficiency to allow greater control over routing, functions that the original design prioritized below survivability. Although IntServ has not achieved widespread acceptance, interest in MPLS appears to be growing.

These developments can be seen as part of a broader move away from viewing routers as static devices that always operate in a particular way and toward looking at the network as a programmable switching fabric that can be reconfigured from store-and-forward routers into virtual circuits as needed. For example, Internet2 (which, as noted earlier, is often held out as proof of the engineering community's conviction that network management is unnecessary) now offers a service that it calls its Interoperable On-demand Network (ION) that allows researchers to establish dedicated point-to-point optical circuits to support large data transfers and other bandwidth-intensive applications. Internet2 notes that the "advanced science and engineering communities . . . are already straining against the limits of today's network capabilities—and capacities" and that advanced media and telepresence applications often need the type of dedicated circuits previously regarded as anathema.[19]

Given the greater flexibility and functionality of today's routers and the increasingly intense demands being placed on them, there seems little reason to require that they always operate in a single, predetermined manner. That said, effective utilization of these new capabilities will doubtlessly require the development of new technical and institutional arrangements. Such innovations and changes may be inevitable if end users are to enjoy the full range of the network's technical capabilities.

---

18 *See* Eric C. Rosen et al., *Multiprotocol Label Switching Architecture*, IETF RFC 3031 (rel. Jan. 2001), *available at* http://www.rfc-editor.org/rfc/pdfrfc/rfc3031.txt.pdf.

19 Internet2, *Internet2 ION* (Sept. 2009), http://www.internet2.edu/pubs/200909-IS-ION.pdf.

G. Low Extra Delay Background Transport (LEDBAT)

Low Extra Delay Background Transport (LEDBAT) is a new IETF congestion management initiative that shows tremendous promise.[20] It is designed to address problems caused by applications that transmit large amounts of data over long periods of time. When this traffic passes through routers that forward on a first in, first out basis without engaging in any active queue management, it imposes heavy delays on all other applications. LEDBAT is designed to address these problems by allowing this high volume, low priority traffic to avoid competing with other best-efforts traffic for its share of the available bandwidth. Instead, LEDBAT permits low priority traffic to step out of the way whenever it encounters any other traffic.

One example of high bandwidth, low priority application would be a service that allows end users to use the Internet to backup their hard disks to remote locations. The end user would likely not care if the service took several hours or even several days. Technologies like LEDBAT permit these end users to run these applications without taking up a disproportionate share of the available capacity or causing network congestion. Peer-to-peer applications similarly generate large amounts of traffic over sustained periods of time. Interestingly, the peer-to-peer community has been very engaged in and largely supportive of the LEDBAT project.

LEDBAT underscores the analytical emptiness of attempting to distinguish between prioritization and degradation. In essence, LEDBAT provides for a level of priority that is worse than best efforts routing. Whether or not it is regarded as degradation depends on what level of service is taken as the relevant baseline. While there is a temptation to regard the current level of service reflected in the current status quo as the natural baseline for comparison, the history of

---

[20] *See Low Extra Delay Background Transport (LEDBAT) Working Group Charter*, INTERNET ENGINEERING TASK FORCE, http://www.ietf.org/html.charters/ledbat-charter.html.

the Internet as well as ongoing debates in the engineering community reveal that there is nothing natural about this level. It is instead simply one of many choices made.

Approaches like LEDBAT reduce the cost of networking by allowing providers to offer higher levels of quality of service without having to expand network capacity. It permits a fairer allocation of bandwidth without requiring all providers to reconfigure their routers to actively manage their queues. It does represent a form of tiered service that will almost certainly involve different levels of pricing. Again, it underscores how permitting such differentials can benefit consumers while simultaneously promoting the goals of the National Broadband Plan.

* * *

This brief overview only touches on a few of innovative ways that providers are deploying specialized services to use bandwidth more efficiently, reduce cost, and provide better service. Providers will need to become even more innovative as the universe of end users continues to become more heterogeneous and as market saturation causes them to focus on delivering greater value to each customer.[21] Most importantly, reducing network costs can help promote the goals of the National Broadband Plan by reducing the number of subscribers needed for an upgrade to the available capacity to break even.

Interestingly, these changes may be just the tip of the iceberg. The U.S. government,[22] the EU,[23] and university-based researchers[24] are pursuing "clean slate" initiatives exploring how

---

21 *See* Christopher S. Yoo, *Product Life Cycle Theory and the Maturation of the Internet*, 104 NW. U. L. REV. (forthcoming 2010).

22 One is the example is DARPA's New Arch initiative. *See* David Clark et al., *New Arch: Future Generation Internet Architecture* 4, 13 (Final Technical Report Dec. 31, 2003), http://www.isi.edu/newarch/iDOCS/final.finalreport.pdf. The National Science Foundation is pursuing initiatives. One is known as the Global Environment for Network Innovations (GENI). *See* National Science Foundation, Global Environment for Network Innovation (GENI), http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501055. Another was originally known as the Future Internet Design (FIND) project. *See* Vint Cerf et al., *FIND Observer Panel Report* (Apr. 9, 2009), http://www.nets-find.net/FIND_report_final.pdf. FIND was subsequently folded into the NSF's Networking Technology and Systems (NeTS) program. *See* National Science Foundation, Networking Technology and Systems (NeTS), http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503307. The NSF's major current initiative is the

the architecture might differ radically if it were designed from scratch today. Within the scope of these projects are even more extensive usage of specialized services to deal with such emerging functionalities as security, mobility, and cloud computing.

II. MOBILE WIRELESS INTERNET ACCESS

Another emerging trend that is transforming U.S. Internet policy is the emergence of wireless as a technological platform for broadband service.[25] The most recent data released by the Commission indicates that wireless already captured nearly 31.2% of the market for high-speed lines as of mid-year 2009, as compared with just over 36.5% for cable modem and 27.0% for ADSL.[26] The expansion of the U.S. wireless broadband market since 2008, the impending deployment of LTE, and the emergence of wireless as the leading broadband platform abroad both suggest that wireless broadband will become increasingly important in the years to come.

Policymakers sometimes suggest that the same principles applying to other broadband technologies should simply be extended to wireless. These suggestions overlook key technological differences between wireless and wireline technologies that policymakers must take into account.

Future Internet Architectures program. *See* National Science Foundation, Future Internet Architectures (FIA), http://www.nsf.gov/news/news_summ.jsp?cntn_id=117611&org=NSF&from=news.

[23] *See, e.g.*, European Commission Directorate-General for the Information Society and Media Project SMART 2008/0049, Internet Development Across the Ages (Jan. 2010), http://cordis.europa.eu/fp7/ict/fire/docs/executive-summary_en.pdf; European Commission, FIRE - Future Internet and Experimentation, (July 6, 2010), http://cordis.europa.eu/fp7/ict/fire/home_en.html.

[24] *See, e.g.*, Jon Crowcroft & Peter Key, *Report from the Clean Slate Network Research Post-SIGCOMM 2006 Workshop*, COMPUTER COMM. REV., Jan. 2007, at 75; Anja Feldmann, *Internet Clean-Slate Design: What and Why?*, COMPUTER & COMM. REV., July 2007, at 59; 100x100 Clean Slate Project, http://100x100network.org/; Stanford University Clean Slate Project, Program Goals, http://cleanslate.stanford.edu/index.php.

[25] These comments are an expanded version of a brief discussion appearing in Christopher S. Yoo, *The Changing Patterns of Internet Usage*, 63 FED. COMM. L.J. (forthcoming December 2010), which was produced with support by Time Warner Cable's Research Program on Digital Communications.

[26] FCC Indus. Analysis & Tech. Div., Wireline Competition Bur., High-Speed Services for Internet Access: Status as of June 30, 2009, at 6 chart 1 (Sept. 2010), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-301294A1.pdf. When measured in terms of advanced service lines, the percentages change to 23.7% for mobile wireless versus 42.3% for cable modem service and 28.8% for ADSL. *Id.* at 7 chart 2. Mobile wireless continues to grow at a much more rapid pace than other technologies.

A. Bandwidth Limits and Local Congestion

Wireless technologies are subject to constraints that are quite different from wireline technologies. As an initial matter, wireless broadband is subject to bandwidth limitations that are much more stringent than those confronted by wireline technologies. While wireless providers can modestly increase capacity by relying on smaller cell sites operating at lower power, they cannot add capacity in the same manner as wireline providers.

In addition, because wireless technologies share bandwidth locally, they are more susceptible to local congestion than many fixed-line services, such as ADSL. These problems are exacerbated by the fact that in wireless networks, data and voice traffic typically share bandwidth, in contrast with telephone and cable companies, which typically place data traffic in a separate channel. Thus, excess data traffic can degrade wireless providers' core business to an extent not possible for other broadband technologies. As a result, many wireless providers rate limit or ban video or peer-to-peer downloads in order to prevent a small number of users from rendering the service completely unusable.[27]

B. The Physics of Wave Propagation

Anyone who has studied physics knows that waves have some unique characteristics. They can reinforce each other in unexpected ways, as demonstrated by unusual echoes audible in some locations in a room and by whispering corners, where the particular shape of the room allows sound to travel from one corner to the other even though a person speaks no louder than a whisper. As noise-reducing headphones demonstrate, waves can also cancel each other out. Waves also vary in the extent to which they can bend around objects and pass through small

[27] A recent, eloquent demonstration of this strategy is the placards aboard the Amtrak Acela express trains asking passengers to refrain from using the WiFi service to download video.

openings, depending on their wavelength. The discussion that follows is necessarily simplified, but is sufficient to convey the intuitions underlying some of the considerations that make wireless networking so complex.

The unique features of waves can cause wireless technologies to face interference problems that are more complex and fast-changing than anything faced by wireline technologies. For example, wireless signals attenuate much more rapidly with distance than do wireline signals, which makes bandwidth much more sensitive small variations in how distant a particular user is from the nearest base station. Moreover, in contrast to wireline technologies, there is an absolute limit to the density of wireless users that can operate in any particular area. Shannon's Law dictates that the maximum rate with which information can be transmitted given limited bandwidth is a function of the signal-to-noise ratio.[28] Unlike wireline transmissions, which travel in a narrow physical channel, wireless signals propagate in all directions and are perceived as noise by other receivers. At some point, the noise becomes so significant that the addition of any additional wireless radios becomes infeasible.

Wireless transmissions also suffer from what are known as *multipath* problems resulting from the fact that terrain and other physical features can create reflections that can cause the same signal to arrive at the same location multiple times. Unless the receiver is able to detect that it is receiving the same signal multiple times, it will perceive multipathing as an increase in the noise floor that reduces the available bandwidth.

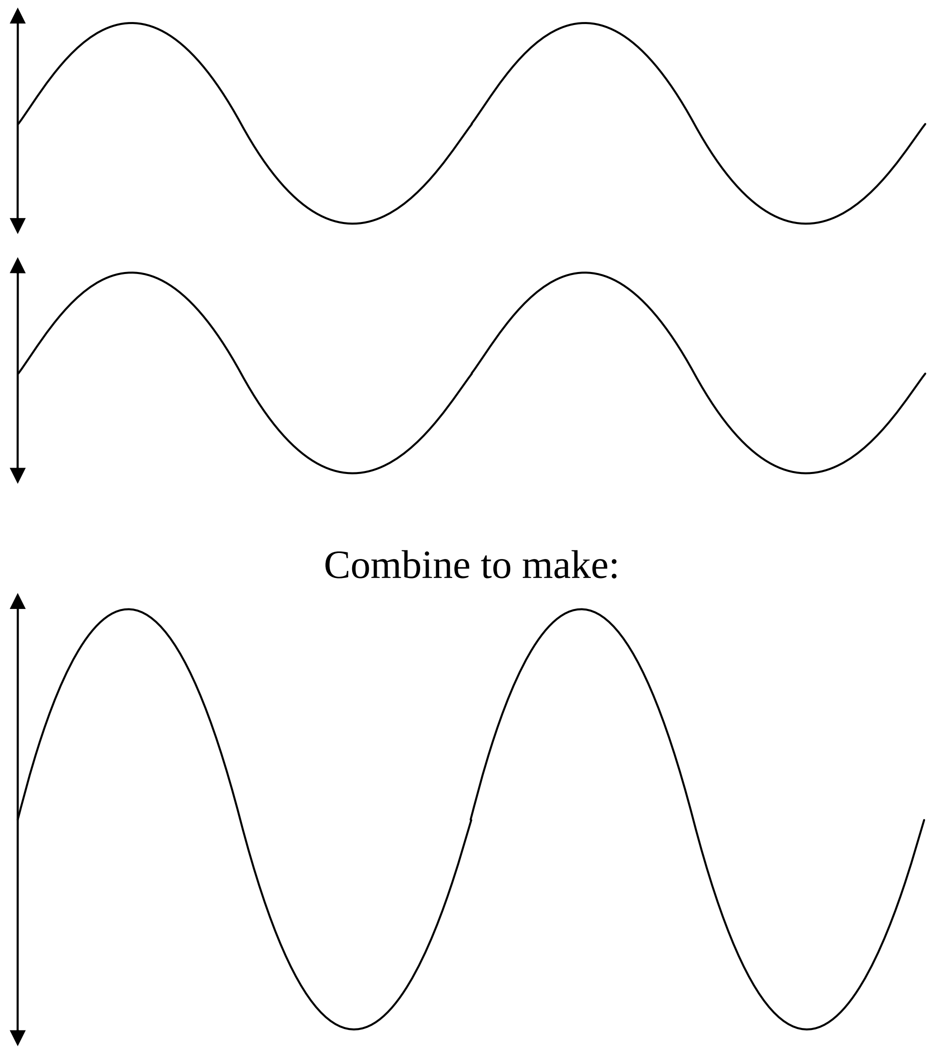
When reflections cause the same signal to arrive by different paths, the signal can arrive either in phase (with the peaks and the valleys of the wave form from the same signal arriving at exactly the same time) or out of phase (with the peaks and the valleys of the wave form from the

---

[28] C.E. Shannon, *A Mathematical Theory of Communication* (pt. 1), 27 BELL SYS. TECH. J. 379 (1948); C.E. Shannon, *A Mathematical Theory of Communication* (pt. 2), 27 BELL SYS. TECH. J. 623 (1948).

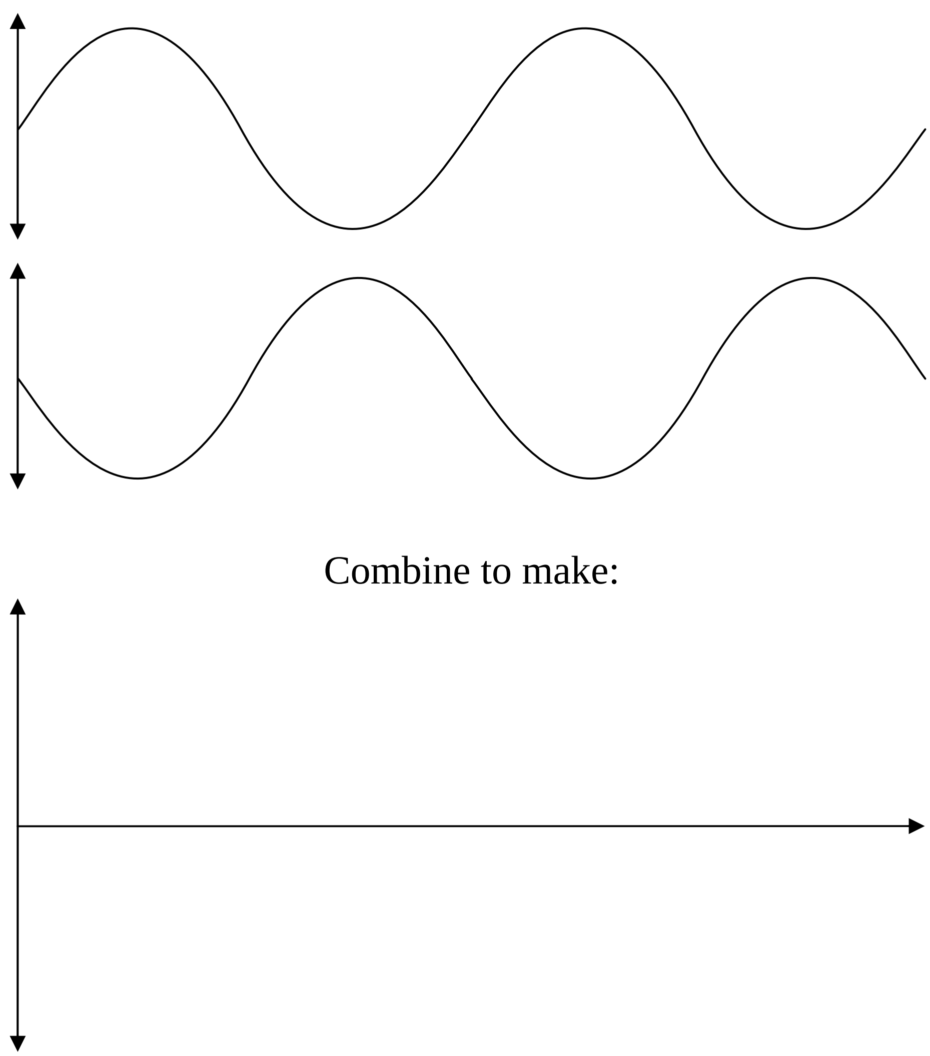
same signal arriving at different times). When waves reflecting off a hard surface arrive in phase, the signal reinforces itself, creating a localized hot spot in which signal is unusually strong.

Figure 1: Reinforcement of Two Wave Forms That Are in Phase



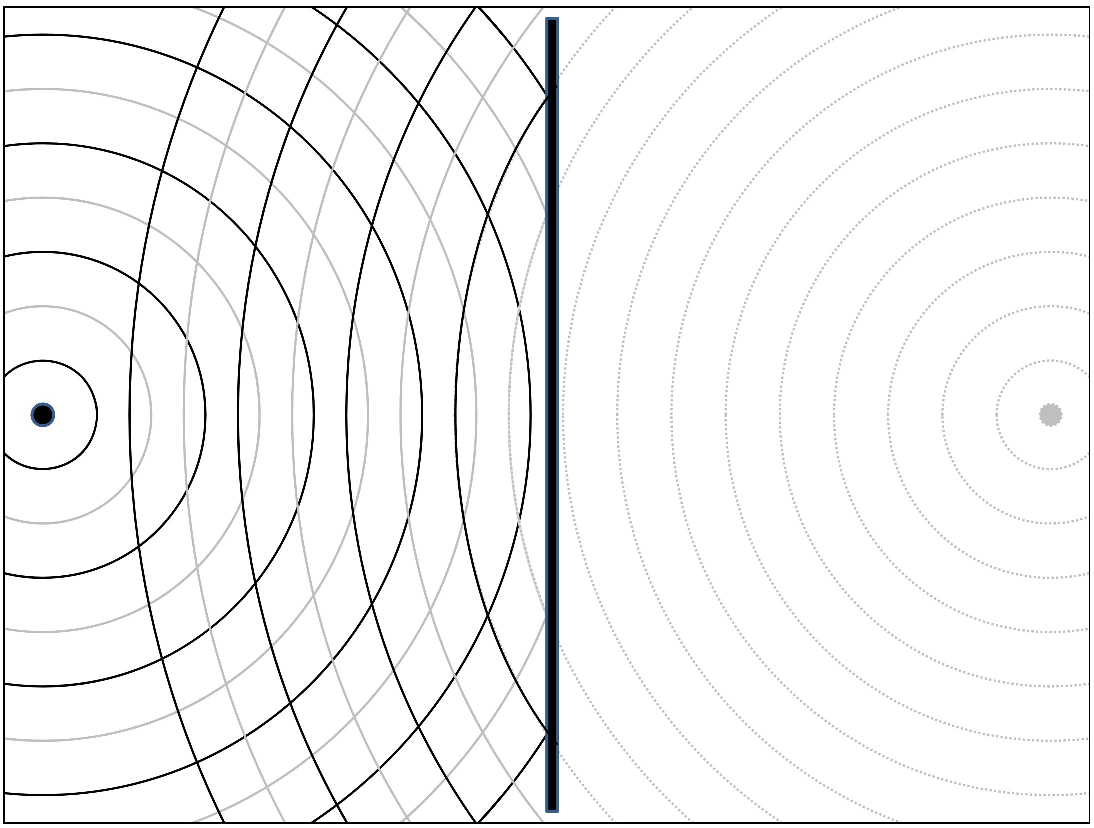
When reflected waves arrive out of phase, they can dampen the signal. When they arrive perfectly out of phase (i.e., 180º out of phase), the reflection can create a dead spot by canceling out the wave altogether. Although smart transmitters and receivers can avoid these problems if they know the exact location of each source and can even use the additional signal to extend the usable transmission range, they cannot do so if the receiver or the other sources are mobile devices whose locations are constantly changing.

Figure 2: Cancellation by Two Wave Forms That Are 180º Out of Phase



A standard result in any physics textbook is that a reflection creates waves that are identical to a point source that is equidistantly located on the other side of the reflective surface. The result is signal strength that is quite unpredictable. Consider the simple diagram in Figure 3, in which that the black circles represent the peaks of the wave form, while the grey circles represent the valleys. The points where two black circles or two grey circles cross represent hot spots where signals reinforce one another. The locations where a black circle crosses a grey circle represent dead spots where waves tend to cancel one another out.

Figure 3: The Problem of Multipath Propagation



Obviously individuals traversing a room might pass through a variety of hot and cold spots. In addition, wave reflections can result not only from immobile objects, such as terrain and buildings, but also from mobile objects, such as cars and trucks. The result is that the amount of bandwidth available can change dynamically on a minute-by-minute basis. A participant at a May 2010 conference held at the University of Pennsylvania related a particularly vivid example of this phenomenon. While living in London, he had an apartment overlooking the famous Speakers' Corner in Hyde Park. Thinking that those in the Speakers Corner might enjoy having WiFi service, he established a WiFi hotspot and pointed a directional antenna at the location only to find that his signal was intermittently blocked even though nothing ever passed directly between his apartment and the Corner. He eventually discovered that the interference arose whenever a double-decker bus was forced to stop at a nearby traffic light. Even though the bus did not directly obstruct with the waves travelling to and from the Speakers' Corner, it created a multipath reflection that periodically cancelled out the direct signal.[29]

For these reasons, many wireless providers implement protocols that dynamically manage their networks based on the available bandwidth, giving priority to time-sensitive applications during times when subscribers are in areas of low bandwidth (such as by holding back email while continuing to provide voice service). They have to do so much more aggressively and dynamically than do wireline providers.

C. Congestion Management

Wireless technologies also require a significantly different approach to congestion management. The primary mechanism for controlling congestion on the Internet was developed in the late 1980s shortly after the Internet underwent a series of congestion collapses. Because congestion is a network-level problem, in many ways the logical solution would have been to address it through a network-level solution, as was done in the original ARPANET, networks

[29] Christian Sandvig, Associate Professor of Communication, University of Illinois, Remarks presented at the Center for Technology, Innovation and Competition's Conference on "Rough Consensus and Running Code: Integrating Engineering Principles into the Internet Policy Debates": How to See Wireless (May 7, 2010). For a description of the project, see PHILIP N. HOWARD, NEW MEDIA CAMPAIGNS AND THE MANAGED CITIZEN xi-xii (2006).

running asynchronous transfer mode (ATM), and many other early corporate networks. However, the router hardware of the time made network-based solutions prohibitively expensive. On the other hand, although edge-based congestion management is feasible, the hosts operating at the edge of the network typically lack the information to know when the network is congested.

Van Jacobson devised an ingenious mechanism by which hosts operating at the edge of the network can infer when the core of the network has become congested.[30] This solution takes advantage of a particular feature of the Transmission Control Protocol (TCP). TCP ensures reliability by requiring the receiving host to send an acknowledgment for every packet that it receives. The sending host estimates the roundtrip time that it would take for a packet and an acknowledgement to traverse the Internet. If the sending host does not receive an acknowledgement within the expected timeframe, it presumes that the packet was lost and resends it.

Jacobson noted that packet loss typically occurs for one of two reasons: (1) transmission errors or (2) discard by a router where congestion has caused its buffer to become full. Because wireline networks rarely dropped packets due to transmission errors, hosts operating at the edge of the network could infer that the failure to receive an acknowledgement within the expected time was a sign of congestion and take this as a signal to reduce congestion by slowing down their sending rates exponentially.[31]

The problem is that this inference is invalid for wireless networks, which drop packets due to transmission error quite frequently, either because of a bad handoff as a mobile user changes cells or because of the interference problems discussed above. When a packet is dropped due to transmission error, reducing the sending rate exponentially is precisely the wrong

---

30 Van Jacobson, *Congestion Avoidance and Control*, 18 COMPUTER & COMM. REV. 314 (1988).

31 *Id.* at 319.

response and would only serve to degrade network performance. Instead, the sending host should resend the dropped packet as quickly as possible without slowing down. In other words, the optimal response for wireless networks may well be the exact opposite of the optimal response for wireline networks.

These differences have caused wireless networks to manage congestion and packet loss in different ways. Some solutions place a "snoop module" at the base station that serves as the gateway used by wireless hosts to connect to the Internet that keeps copies of all packets that are transmitted and monitors acknowledgments passing in the other direction. When the base station detects that a packet has failed to reach a wireless host, it resends the packet locally instead of having the sending host do so.[32] Other solutions call for the sending host to be aware of when its transmission is carried in part by a wireless link and to distinguish between losses due to congestion and losses due to transmission errors. Still other solutions call for a split connection, in which the sending host establishes one TCP connection with an IP gateway in the middle of the network where the transmission shifts to wireless and a separate TCP connection between the IP gateway and the receiving host.[33] Some of these solutions violate the semantics of IP. All of them require introducing traffic management functions into the core of the network to a greater extent than originally envisioned by the Internet's designers.

D. The Heterogeneity of Devices

Wireless technologies do not vary only in terms of transmission technologies. They also vary in terms of end-user devices. Instead of relying on a personal computer, wireless broadband subscribers connect to the network through a wide variety of smart phones. These devices are

---

[32] Hari Balakrishnan et al., *Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks*, 1 WIRELESS NETWORKS 469 (1995).

[33] JAMES F. KUROSE & KEITH W. ROSS, COMPUTER NETWORKING: A TOP-DOWN APPROACH 585-86 (5th ed. 2010); ANDREW S. TANENBAUM, COMPUTER NETWORKS 553-54 (4th ed. 2003).

much more sensitive to power consumption than are PCs, which sometimes leads wireless network providers to disable certain functions that shorten battery life to unacceptable levels, for example because they either employ analog transmission or search constantly for an available connection. In addition, wireless devices have much less processing capacity and employ less robust operating systems that do the laptop and personal computers typically connected to wireline services. As a result, they are more sensitive to conflicts generated by multiple applications, which can cause providers to be much more careful about which applications to permit to run on them.

Wireless devices also tend to be much more heterogeneous in terms of operating systems and input interfaces (including keyboards and touch screens). As a result, the dimensions and levels of functionality offered by particular wireless devices vary widely. It seems too early to predict with any confidence which platform or platforms will prevail. Furthermore, as noted earlier, many wireless networks address bandwidth scarcity by giving a higher priority to time-sensitive applications, which typically requires close integration between network and device. These features underscore the extent to which variations in particular devices are often an inextricable part of the functionality of the network.[34]

Even more fundamentally, wireless devices interconnect with the network on very different terms. Devices connected to wireline networks have IP addresses that are visible to all other Internet-connected hosts. Wireless devices, in contrast, do not have IP addresses. Instead, Internet connectivity is provided by an IP gateway located in the middle of the network that connects to individual wireless devices using a telephone-based technology rather than IP. Stated in technical terms, wireless broadband devices operate at Layer 2 rather than Layer 3 of

---

[34] Charles L. Jackson, *Wireless Efficiency versus Net Neutrality*, 63 FED. COMM. L.J. (forthcoming March 2011).

the Internet protocol stack. This means that all current wireless devices do not have the end-to-end visibility enjoyed by true Internet-enabled devices. They also necessarily depend on a virtual circuit between the Internet gateway and the wireless device. Wireless devices will eventually connect through the Internet protocol once fourth-generation wireless technologies such as LTE are deployed. Until that time, wireless devices necessarily will connect to the Internet on different and less open terms than devices connected through wireline networks.

E. Routing and Addressing

Another problem confronting wireless broadband results from the fact that the Internet developed at a time when computers did not move. As a result, the architecture could use a single address to specify both the identity of a particular host as well as where that host was connected to the network. The advent of mobility has caused the unity of identity and location to break down. A single mobile device may now connect to the network through any number of locations. Although the network could constantly update the routing table to reflect the host's current location, doing so would require propagating the updated information to every router in the network as well as an unacceptably large number of programs and databases.

Instead, mobile devices typically designate a router on its home network that has a fixed, permanent IP address as a "home agent" that serves as the initial contact point for all IP-based communications. Anyone seeking to contact the host would first send the packets to the home agent, which would then encapsulate the packets in another packet and forward them to wherever the mobile host is currently located. Managing mobile communications in this manner is surprising complex and requires protocols for home agents to notify others of its location and to encapsulate traffic bound for the mobile host and for mobile hosts to register and deregister their current location with their home agents, notify the foreign network that they are currently attached to them, and decapusulate the packets they receive. Sending communications via the home agent also suffers from the inefficiency of what is sometimes called "triangle routing," because instead of passing directly from the sending host to the receiving host, traffic must travel first from the sending host to the home agent and then from the home agent to the receiving host. In the extreme case, a communication between two mobile hosts located next to one another in a conference room on the west coast might have to travel back and forth across the country if one of them has a home agent located on the east coast. The home agent can eliminate triangle routing by passing the mobile host's current location on to the sender so that the sender may forward subsequent packets to it directly. The initial communications must still bear the inefficiency of triangle routing. Moreover, such solutions become much more difficult to implement if the mobile agent is constantly on the move.[35]

As a separate matter, wireless technologies are also causing pressure on the way the amount of resources that the network must spend on keeping track of Internet addresses. Tier 1 ISPs necessarily must maintain complete routing tables that identify the IP address for every host connected to the Internet. The current system relies on route aggregation to keep routing tables from growing out of control. This mechanism can be illustrated by analogy to the telephone system. Consider a party in Los Angeles who is attempting to call the main telephone number for the University of Pennsylvania, which is (215) 898-5000. So long as all calls to the 215 area code pass through the same outbound link, a phone switch in Los Angeles could represent all telephone numbers in that area code with a single entry in its routing table. Similarly, so long as all telephone numbers in the 898 directory are connected to the same central office, switches within Philadelphia need not maintain separate entries for each phone number in that directory.

---

[35] DOUGLAS E. COMER, INTERNETWORKING WITH TCP/IP: PRINCIPLES, PROTOCOLS, AND ARCHITECTURE 339-46 (5th ed. 2006); KUROSE & ROSS, *supra* note 33, at 566-77; TANENBAUM, *supra* note 33, at 372-75, 462-64.

Instead, they can represent all telephone numbers located in (215) 898-xxxx with a single entry. The Internet employs a system known as Classless InterDomain Routing (CIDR) to aggregate routes. CIDR is even more flexible. It can aggregate routes at any number of digits rather than being limited in the manner of area codes and directories with the digits of course being represented in binary. This strategy depends on the address space remaining compact. In other words, this approach will fail if the 215 area code includes phone numbers that are not located in Philadelphia. If that is the case, the routing table will have to use separate entries to keep track of every single address. The problem is that true mobile addressing fragments the geographic compactness of the address space.

Another problem is somewhat more subtle. The current architecture is built on the implicit assumption that Internet addresses change on a slower timescale than do communication sessions. So long as the address architecture changes at a slower timescale, any particular Internet-based communication may take the address architecture as given. Mobility, however, increases the rate at which the address architecture changes. In addition, because addressing is handled on a decentralized basis, information about changes in the address architecture takes time to spread across the Internet. Increases in the rate with which the address space changes can cause communications sessions to fail and create the need for a new way to manage addresses.

Others have proposed radical changes in the addressing and routing architecture. One approach would replace the single address now employed in the network with two addresses: one to identify the particular machine and the other to identify its location. Whatever solution is adopted would represent a fundamental change in the network layer than unifies the entire Internet.

* * *

The net result is that mobile wireless broadband networks operate on principles that are quite different from those governing the rest of the Internet. Bandwidth limitations require that wireless providers manage their networks more intensively than those operating networks based on other technologies. The fact that smartphones do not have IP addresses and the higher incidence of packet loss require that wireless networks employ virtual circuits and embed intelligence in the network to provide Internet access and to handle the problems of congestion. The unpredictability of signal strength resulting from the physics of wave propagation can necessitate more extensive supervision than other technologies require, as do the realities of system conflicts and power consumption. Lastly, mobility is placing pressure on the routing and addressing space that may soon require more fundamental changes. The industry has not yet reached consensus on the best approach for addressing all of these concerns. In its consideration of regulatory interventions, the Commission must be careful to create a regime that takes these differences into account.

III. POLICY IMPLICATIONS

Mobile wireless Internet access services thus raise challenges with which the original architecture, which consisted primarily of stationary devices connected to the telephone system, did not have to contend. The potential benefits of network management and specialized services are also myriad. Moreover, the current pace of technological innovation guarantees that new challenges and new solutions will continue to appear on the scene. It is for this reason that I have long advocated an ex post, case-by-case approach to evaluating claims of allegedly improper behavior[36] and encouraged to see this approach emerge as a source of consensus.[37]

---

36 Christopher S. Yoo, *Would Mandating Broadband Network Neutrality Help or Hurt Competition? A Comment on the End-to-End Debate*, 3 J. ON TELECOMM. & HIGH TECH. L. 23, 44-47, 58-59 (2004) [hereinafter

The primary question is whether an ex post, case-by-case regimes can provide sufficient predictability to support innovation and investment. As an initial matter, it bears mentioning that many key areas of U.S. law (such as torts, contracts for services, and property) are governed by case-by-case, common law decisionmaking and yet still manage to provide enterprises and innovators with enough certainty to pursue their initiatives.

That said, it is unquestionably true that it would be beneficial for the Commission to provide advance guidance to those potentially subject to this new regulatory regime. Since even proponents of the open Internet acknowledge that it is primarily about vertical foreclosure made possible by centralized control over both content/applications and conduit,[38] the most logical starting point for analysis is the set of principles developed by the courts and the antitrust enforcement agencies to govern vertical integration[39] and vertical restraints (such as an exclusive dealing or tying contract) that achieve what amounts to vertical integration via contract rather than merger.[40]

Yoo, *Comment on End-to-End*]; Yoo, *supra* note 2, at 7-8, 24, 75; Christopher S. Yoo, *Network Neutrality and the Economics of Congestion*, 94 GEO. L.J. 1847, 1854-55, 1900, 1908 (2006) [hereinafter Yoo, *Economics of Congestion*]; Christopher S. Yoo, *What Can Antitrust Contribute to the Network Neutrality Debate?*, 1 INT'L J. COMM. 493, 504, 510, 530 (2007) [hereinafter Yoo, *What Can Antitrust Contribute*]; Christopher S. Yoo, *Network Neutrality, Consumers, and Innovation*, 2008 U. CHI. LEGAL F. 179, 186-87, 212, 227, 238, 246-47, 257, 261 [hereinafter Yoo, *Consumers, and Innovation*]; Christopher S. Yoo, *Network Neutrality after* Comcast*: Toward a Case-by-Case Approach to Reasonable Network Management, in* NEW DIRECTIONS IN COMMUNICATIONS POLICY 55, 56-57, 81-83 (Randolph J. May ed., 2009).

37 Further Inquiry into Two Under-Developed Issues in the Open Internet Proceeding 1, DA 10-1667 (F.C.C. Sept. 1, 2010), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-10-1667A1.pdf.

38 LAWRENCE LESSIG, THE FUTURE OF IDEAS 165-66 (2001); Daniel L. Rubinfeld & Hal J. Singer, *Open Access to Broadband Networks: A Case Study of the AOL/Time Warner Merger*, 16 BERKELEY TECH. L.J. 631, 632-34 (2001); Tim Wu, *The Broadband Debate, A User's Guide*, 3 J. ON TELECOMM. & HIGH TECH. L. 69, 84-85 (2004); *see also* 3A PHILIP E. AREEDA & HERBERT HOVENKAMP, ANTITRUST LAW ¶ 771a, at 192-94 (3d ed. 2008) (pointing out that all questions about access to essential facilities are necessarily questions about vertical integration).

39 U.S. Department of Justice & Federal Trade Commission, Non-Horizontal Merger Guidelines, 57 Fed. Reg. 41,552 (1992), *available at* http://www.justice.gov/atr/public/guidelines/2614.htm [hereinafter Non-Horizontal Merger Guidelines].

40 *See, e.g.*, *Jefferson Parrish Hosp. Dist. No. 2 v. Hyde*, 466 U.S. 2, 15-16 (1984).

A. A General Framework for Analysis

Established principles of antitrust law establish a number of preconditions that must be met before a decisionmaker should seriously entertain any claim of anticompetitive conduct. First, the relevant markets must be structured to give the actor charged with anticompetitive conduct the *ability* to act in an anticompetitive manner. This typically requires a high level of concentration in the primary market that the actor is allegedly using as leverage.[41] In addition, the secondary market that is allegedly being rendered uncompetitive must be concentrated[42] and protected by entry barriers.[43]

Second, the actor charged with anticompetitive conduct must have the *incentive* to engage in the allegedly anticompetitive conduct.[44] Even if the actor has the ability to engage in the conduct in question, regulators need not intervene in the absence of proof that the conduct would be profitable.

Third, the conduct in question must actually harm consumers, in that whatever anticompetitive effects that may exist are not offset by any countervailing *efficiencies*.[45] For example, there are four classic sources of efficiencies recognized by every textbook on industrial organization.[46] First, it is universally recognized that integration between successive

---

41 *See* Non-Horizontal Merger Guidelines, *supra* note 39, §§ 4.131, 4.213 (generally requiring HHI to exceed 1800 before a vertical merger would raise anticompetitive concerns); *Jefferson Parrish*, 466 U.S. at 13-14 (requiring proof of market power in the primary market).

42 *See* Non-Horizontal Merger Guidelines, *supra* note 39, § 4.211 (noting impropriety of challenging vertical mergers when sufficient unintegrated capacity exists); *Jefferson Parrish*, 466 U.S. at 16 (requiring proof of substantial foreclosure of the secondary market); *Barry Wright Corp. v. ITT Grinnell Corp.*, 724 F.2d 227, 237 (1st Cir. 1983) (Breyer, J.) (ruling that legality of exclusive dealing contract depends on severity of foreclosure).

43 *See* Non-Horizontal Merger Guidelines, *supra* note 39, § 4.212 (noting impropriety of challenging vertical mergers when entry into the secondary market is easy).

44 *See, e.g.*, *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574, 5587-9594 (1986).

45 Non-Horizontal Merger Guidelines, *supra* note 39, §§ 4.135, 4.24.

46 *See* ROGERS D. BLAIR & DAVID L. KASERMAN, LAW AND ECONOMICS OF VERTICAL INTEGRATION AND CONTROL 18-23, 31-42, 48-52 (1983); F.M. SCHERER & DAVID ROSS, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE 519-27, 551-55 (3d ed. 1990); JEAN TIROLE, THE THEORY OF INDUSTRIAL ORGANIZATION 174-81 (1988).

monopolists would not only be profitable, but would also benefit consumers by eliminating the classic problem of double marginalization.[47] In some (but not all) cases, vertical integration can also rationalize input substitution when inputs can be used in variable proportions.[48] Vertical integration can also eliminate free riding on presale services.[49] Finally, as Oliver Williamson recognized in the seminal work for which he was recently awarded the Nobel Prize, vertical integration can also benefit consumers by eliminating the transaction costs needed to guard against opportunistic behavior.[50]

Fourth, the proposed remedies must be tailored to specifically address the particular anticompetitive conduct charged.[51] Each of these preconditions must be evaluated in light of the particular theory of anticompetitive harm being alleged. Indeed, one of the most trenchant criticisms of the government's case against Microsoft is the disconnect between the theory alleged, the evidence adduced, and the remedy the government sought to impose.[52]

Determining liability is complicated by the fact that the game theoretic models that dominate the current literature on industrial organization are highly stylized and are highly sensitive to small changes in theoretical and empirical assumptions.[53] For example, leading

---

47 For the seminal statements, see Joseph J. Spenlger, *Vertical Integration and Antitrust Policy*, 58 J. POL. ECON. 347 (1950); Fritz Machlup & Martha Taber, *Bilateral Monopoly, Successive Monopoly, and Vertical Integration*, 27 ECONOMICA 101 (1960). For an overview, see Christopher S. Yoo, *Vertical Integration and Media Regulation in the New Economy*, 19 YALE J. REG. 171, 192-93 (2002).

48 For the seminal economic analyses, see L.W. McKenzie, *Ideal Output and the Interdependence of Firms*, 61 ECON. J. 785 (1951); Meyer Burstein, *A Theory of Full-Line Forcing*, 55 NW. U. L. REV. 62 (1960); and John M. Vernon & Daniel A. Graham, *Profitability of Monopolization by Vertical Integration*, 79 J. POL. ECON. 924 (1971). For an overview, see Yoo, *supra* note 47, at 189-90.

49 For the seminal analysis, see Lester G. Telser, *Why Should Manufacturers Want Fair Trade?*, 3 J.L. & ECON. 86 (1960). For an overview, see Yoo, *supra* note 47, at 195-96. The Supreme Court explicitly embraced this analysis in *Continental TV Inc. v. GTE Sylvania, Inc.*, 433 U.S. 36, 55 (1976).

50 *See* OLIVER E. WILLIAMSON, MARKETS & HIERARCHIES (1975). For an overview, see Yoo, *supra* note 47, at 193-98.

51 U.S. DEPARTMENT OF JUSTICE, COMPETITION AND MONOPOLY: SINGLE FIRM CONDUCT UNDER SECTION 2 OF THE SHERMAN ACT 143 (2008).

52 *See* Timothy J. Brennan, *Do Easy Cases Make Bad Law? Antitrust Innovations or Missed Opportunities in* United States v. Microsoft, 69 GEO. WASH. L. REV. 1042 (2001).

53 James C. Cooper et al, *Vertical Antitrust Policy as a Problem of Inference*, 23 INT'L J. INDUS. ORG. 639, 648 (2005); Brennan; Hovenkamp.

antitrust scholar Herbert Hovenkamp concludes that post-Chicago theories of vertical integration "may not be quite ready for prime time," noting that "while many post-Chicago school theories admit a greater range of anticompetitive strategic behaviors, most of these occur only under strictly defined circumstances. If these circumstances do not exist, then the contemplated strategy will not work."[54] Hovenkamp cautions that decisionmakers must take great care to ensure that complainants actually adduce the evidence needed to establish their claim.[55]

A paper authored by several members of the senior staff of the Federal Trade Commission draws similar conclusions, pointing out that the models in the current theoretical literature suggest that vertical integration may benefit as well as harm consumers.[56] Moreover, the results are fragile and often collapse whenever any of the models' assumptions are relaxed.[57] Because of this, a recent paper by a scholar who would later serve as the Antitrust Division's Chief Economist cautioned against drawing strong policy inferences without verifying that the conditions for anticompetitive effects are actually met and not offset by countervailing efficiencies.[58] This is particularly difficult because the structural preconditions needed to give make anticompetitive effects feasible are also the structural conditions under which vertical integration tends to create efficiencies.[59] The author of a leading industrial organization textbook similarly concludes, because "most vertical restraints can increase or decrease welfare, depending on the environment," "[t]heoretically, the only defensible position on vertical restraints seems to be the rule of reason."[60]

---

[54] Herbert Hovenkamp, *Post-Chicago Antitrust: A Critique and Review*, 2001 COLUM. BUS. L. REV. 257, 278, 326.

[55] *Id.* at 279.

[56] Cooper, *supra* note 53, at 643-44, 646-47.

[57] *Id.* at 644.

[58] *Id.* at 646-47.

[59] *Id.* at 647, 658.

[60] TIROLE, *supra* note 46, at 186.

The stylized nature of the theories provides a nice demonstration of the distinction between what economist Franklin Fisher has called "exemplifying theory" and "generalizing theory" in analyzing public policy.[61] Generalizing theory relies on fairly general assumptions to establish broad propositions that apply under a wide range of circumstances. Exemplifying theory, in contrast, employs specialized assumptions to show what can happen under particular circumstances.[62] The specificity of exemplifying theory can play an important role in isolating the effect of particular economic considerations or in serving as possibility theorems demonstrating the potential existence of particular phenomena. As such, exemplifying theory is very helpful in establishing what *could* happen. Without empirical validation, its sensitivity to small changes in assumptions make it poorly suited for identifying what *will* happen.[63]

The importance of only intervening only when the empirical evidence clearly shows a likelihood of harm to consumers is further underscored by the growing body of empirical evidence showing that vertical integration and vertical restraints tend to benefit consumers in the vast majority of cases. For example, a recent survey of the literature by leading vertical integration theorist and former FCC Chief Economist Michael Riordan concludes, "A general presumption that vertical integration is pro-competitive is warranted by a substantial economics literature identifying efficiency benefits of vertical integration, including empirical studies demonstrating positive effects of vertical integration in various industries."[64]

A survey of the empirical literature published in the same volume reviewed twenty-three published empirical studies of vertical restraints. Despite the relatively small sample size, the

---

61 Franklin M. Fisher, *Games Economists Play: A Noncooperative View*, 20 RAND J. ECON. 113, 117 (1989).

62 Fisher, *supra* note 61, at 117-18.

63 *Id.* a 118; *see also* Alan J. Meese, *Tying Meets the New Institutional Economics: Farewell to the Chimera of Forcing*, 146 U. PA. L. REV. 1, 89 (1997) ("Per se rules cannot be established by exemplifying theories.").

64 Michael H. Riordan, *Competitive Effects of Vertical Integration, in* HANDBOOK OF ANTITRUST ECONOMICS 145, 169 (Paolo Buccirossi ed., 2008).

authors found the empirical evidence to be “quite striking,” “surprisingly consistent,” “consistent and convincing,” and even “compelling.”[65] As a general matter, “privately imposed vertical restraints benefit consumers or at least do not harm them,” while government mandates or prohibitions of vertical restraints “systematically reduce consumer welfare or at least do not improve it.”[66] Together “[t]he evidence . . . supports the conclusion that in these markets, manufacturer and consumer interests are apt to be aligned, while interference in the market [by the government] is accomplished at the expense of consumers (and of course manufacturers).”[67] The authors conclude that “the empirical evidence suggests that in fact a relaxed antitrust attitude towards [vertical] restraints may well be warranted.”[68]

Another recent study conducted by four members of the FTC’s staff surveying twenty-two published empirical studies found “a paucity of support for the proposition that vertical restraints/vertical integration are likely to harm consumers.”[69] Indeed, only one study unambiguously found that vertical integration harmed consumers, and “in this instance, the losses are miniscule ($0.60 per cable subscriber per year).”[70] On the other hand, “a far greater number of studies found that the use of vertical restraints in the particular context studied improved welfare unambiguously.”[71] The survey thus concluded that “[m]ost studies find evidence that vertical restraints/vertical integration are pro-competitive.”[72] The weight of the

---

65 Francine Lafontaine and Margaret Slade, *Exclusive Contracts and Vertical Restraints: Empirical Evidence and Public Policy*, *in* HANDBOOK OF ANTITRUST ECONOMICS, *supra* note 64, at 392, 408, 409.
66 *Id.* at 408.
67 *Id.* at 409.
68 *Id*.
69 Cooper, *supra* note 53, at 648.
70 *Id.*
71 *Id.*
72 *Id.* at 658.

evidence thus "suggests that vertical restraints are likely to be benign or welfare enhancing,"[73] which in turn provides empirical support for placing the burden on those opposing the practice.[74]

In the absence of specific evidence of harm to consumers in a particular case with respect to particular theory, the empirical evidence that exists provides no basis for viewing vertical coordination as presumptively suspect. This is particular true because when a particular practice is novel, the data needed to evaluate that practice's economic impact may not exist. Placing the burden of proof would prevent ambiguous cases from going forward, which would eliminate the breathing room that new practices need in order to come into being. Moreover, economic strategies are often not the result of systematic planning based on identified benefits, but rather on a process of trial-and-error in which firms experiment with new ways of doing business without always having a clear idea of the precise nature of the benefits.[75]

These considerations counsel strongly in favor of placing the burden of proof on those alleging that a particular practice would harm consumers. Although such reticence might be appropriate if the consequences were catastrophic or irreversible,[76] the temporary reductions in innovation that are feared do not rise to the level usually required to meet those standards.

B. Application of the Framework to Michael Whinston's Tying Model

Applying my proposed analysis to Michael Whinston's seminal tying model illustrates how it would work in fact.[77] Whinston's analysis represents perhaps the most important recent contribution to the theory of vertical restraints and has been invoked by network neutrality

---

73 *Id.* at 662.
74 *Id.* at 661-62.
75 Armen Alchian, *Uncertainty, Evolution, and Economic Theory*, 58 J. POL. ECON. 211 (1950).
76 Yoo, *supra* note 2, at 74; Yoo, *Economics of Congestion*, *supra* note 36, at 1851-52, 1855, 1899, 1908.
77 Michael D. Whinston, *Tying, Foreclosure, and Exclusion*, 80 AM. ECON. REV. 837 (1990). The following analysis is adapted from Yoo, *Consumers, and Innovation*, *supra* note 36, at 249-56.

proponents in support of their arguments.[78] Whinston effectively rebutted the Chicago School's calls to treat tying as legal per se[79] by showing the existence of circumstances under which tying can be profitable and can exclude competitors.

Whinston's model turns on three key factual assumptions. First, it assumes that the firm engaging in the exclusionary conduct is a monopolist in one market, called the primary market, *A*, which consists of a_m potential buyers. Second, it assumes the existence of a market for a complementary product, *B*. Although everyone who purchases *A* also purchases *B*, there are also consumers who purchase only *B*, but not *A*. Third, the market for *B* must be characterized by economies of scale and that the minimum efficient scale for *B* must be larger than the number of *B* consumers who do not purchase *A*.

If the *A* monopolist does not tie its product to *B*, it can only exercise monopoly power over the number of consumers who purchase $A(a_m)$. If, on the other hand, the *A* monopolist forces all *B* consumers who also purchase $A(b_m)$ to purchase from it by tying *A* to *B*, it can increase its profits by foreclosing the market for *B* and earning monopoly profits on a quantity that includes $a_m + b_i$. If the minimum efficient scale (s) is larger than the number of consumers who buy only $B(b_i)$, any independent producer of *B* will operate at a cost disadvantage versus the *A* monopolist and will not be able to compete. This in turn allows the *A* monopolist to drive all independent producers of *B* out of the market and to exercise market power over all *B* consumers (including both b_m and b_i) and not just those who also purchase $A(b_m)$. In this manner, the *A*

---

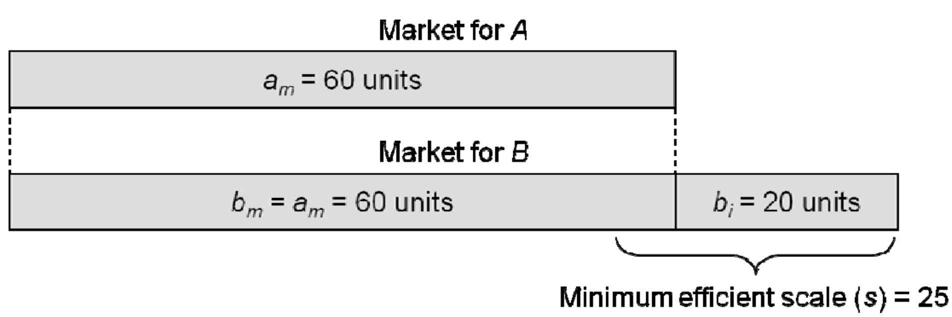
78 *See* Barbara van Schewick, *Towards an Economic Framework for Network Neutrality Regulation*, 5 J. ON TELECOMM. & HIGH TECH. L. 329, 342-52 (2007); and Brett Frischmann & Barbara van Schewick, *Net Neutrality and the Economics of the Information Superhighway: A Reply to Professor Yoo*, 47 JURIMETRICS J 383, 412-14 (2007).

79 *See, e.g.*, ROBERT H. BORK, THE ANTITRUST PARADOX 380 (1978) ("[A] logically consistent law would have to accept the legality of all tying arrangements. . . .").

monopolist can exclude rival producers of *B* and thereby increase the number of consumers over which it can exert monopoly power.

The following numerical example may help illustrate the intuitions underlying Whinston's model. Suppose that the market for good *A* is 60 units and that that market is dominated by a monopolist. Suppose further that all 60 people who purchase good *A* also purchase good *B*, and there are 20 other consumers who only purchase good *B*. Whinston pointed out that if minimum efficient scale for producing *B* is greater than 20 units (say 25 units), the *A* monopolist can obtain a monopoly in the market for *B* simply by tying *A* and *B* together. This is because tying *A* and *B* only leaves 20 uncommitted consumers of *B*. Because the minimum efficient scale for producing *B* is greater than 20, every other potential producer of *B* will operate at a cost disadvantage in *B* production vis-à-vis the monopolist and thus will be unable to compete effectively. This allows the producer of *A* to exercise monopoly power over not only the 60 consumers who purchase both *A* and *B*, but also the 20 additional consumers who purchase only *B*.

Figure 4: Whinston's Tying Model – The Monopoly Case



A close analysis of the model reveals how much its results depend on its assumptions. Consider first the fact that the model depends on a very precise relationship between the number of consumers who purchase only $B(b_i)$ and the minimum efficient scale for producing $B(s)$. The model strictly requires that $s > b_i$. This is because the *A* monopolist can only tie up those *B* consumers who also purchase *A*. If the number of consumers that purchase only *B* exceeds the minimum efficient scale for producing *B* (that is, if $s < b_i$), a rival producer of *B* will not operate at a cost disadvantage and will not be driven from the market. Thus, Whinston's model in effect requires that the number of consumers who purchase only $B(b_i)$ be relatively small and that the minimum efficient scale for producing $B(s)$ be relatively large. Any deviation from this relationship prevents the effect that Whinston has identified from materializing.

In the context of network neutrality, the primary market is the market for last-mile Internet access, and the secondary market is the market for content and applications. This raises the empirical questions of (1) whether there are consumers of Internet content and applications who do not also purchase Internet access services and (2) how large is the minimum efficient scale for providing Internet content and applications. Note also that the number of consumers of Internet content and applications who do not also purchase Internet access services must be small relative to minimum efficient scale; otherwise the effect that Whinston identifies will not arise.

Network neutrality proponents assert that the first condition is met (that is, there are consumers of Internet content and applications who do not also purchase Internet access services) because the market for Internet access is local. For example, the relevant market for last-mile Internet services to my home is the city of Philadelphia, which for the time being we shall assume is a monopoly. The existence of consumers in other cities outside of Philadelphia represents consumers of the secondary good (Internet content and applications) who do not also consume the primary good (last-mile Internet access in Philadelphia).

I have long suggested that the relevant market for Internet content and applications is national, not local. As I noted elsewhere:

> Major web-based providers, such as Amazon.com or eBay, are focused more on the total number of customers they are able to reach nationwide than they are on their ability to reach customers located in any specific metropolitan area. The fact that they may be unable to reach certain customers is of no greater concern, however, than the fact that manufacturers of particular brands of cars, shoes, or other conventional goods are not always able to gain distribution in all parts of the country. . . . The proper question is thus not whether the broadband transport provider wields market power vis-à-vis broadband users in any particular city, but rather whether that provider has market power in the national market for obtaining broadband content.[80]

Indeed, the FCC and the D.C. Circuit relied on precisely this reasoning when ruling that the geographic scope of the market in which video programmers bargain with multichannel video providers is national rather than local. As the FCC and the D.C. Circuit both recognized, a television programmer's viability does not depend on its ability to reach viewers in any particular localities, but rather on the total number of viewers it is able to reach nationwide. So long as a cable network can reach a sufficient number of viewers to ensure viability, the fact that a particular network owner may refuse carriage in any particular locality is of no consequence.[81] The FCC has similarly rejected the notion that the local market power enjoyed by early cellular telephone providers posed any threat to the cellular telephone equipment market, since any one cellular provider represented a tiny fraction of the national equipment market.[82]

Simply put, it is national reach, not local reach, that matters. This in turn implies that the relevant geographic market is a national one, not a local one. What matters is not the percentage of broadband subscribers that any particular provider controls in any geographic area, but rather the percentage of a nationwide pool of subscribers that that provider controls.

---

80 Yoo, *supra* note 2, at 72-73; *accord* Yoo, *Economics of Congestion*, *supra* note 36, at 1892-93; Yoo, *Comment on End-to-End*, *supra* note 36, at 52; Yoo, *supra* note 47, at 254.

81 *See Time Warner Entm't Co. v. FCC*, 240 F.3d 1126, 1131-32 (D.C. Cir. 2001) (citing *Implementation of Section 11(c) of the Cable Television Consumer Protection and Competition Act of 1992*, Third Report and Order, 14 FCC Rcd. 19098, 19114-18 ¶¶ 40-50 (1999)).

82 *See Bundling of Cellular Customer Premises Equipment and Cellular Service*, Report and Order, 7 FCC Rcd. 4028, 4029-30 ¶ 13 (1992).

Some network neutrality proponents assert that customers' interest in local news and local yellow pages is enough to render the relevant geographic market local.[83] Although that observation is true as far as it goes, acknowledging that the proper geographic scope for a small number websites is local does not undercut the fact that the relevant geographic market for the vast majority of websites and applications is national rather than local.[84] If so, it is not clear whether there is any residuum of purchasers of *B* (Internet content and applications) who do not also purchase *A* (last-mile Internet services). Even so, it would not support the type of general access rule that network neutrality proponents seek. At most, it would support a remedy granting access to local content that would not apply to national content.

But even if one were to accept the assertions of network neutrality proponents and assume that the relevant market is local, that would render the Whinston model inapposite. This is because leverage under the Whinston model is impossible if the number of consumers who purchase *B* without also purchasing *A* is relatively small. If not, the large number of uncommitted consumers of *B* makes it easy for independent purchasers of *B* to achieve minimum efficient scale. If the relevant market is local, the number of consumers who purchase *B* without purchasing *A* (in the example above, the number of consumers of Internet content and applications outside of Philadelphia) will be very large, which renders it increasingly unlikely that this will exceed the economies of scale.

Consider next the assumption that the market for *B* (Internet content and applications) is characterized by large economies of scale. The fact that most content and applications markets have not collapsed into natural monopolies suggests either that the scale economies created by

[83] Frischmann & von Schewick, *supra* note 78, at 412-16.

[84] See Christopher S. Yoo, *Rethinking the Commitment to Free, Local Television*, 52 EMORY L.J. 1579, 1657-58 (2003) (making a similar point in the context of local-interest vs. national-interest television).

fixed costs are particularly large[85] or that there is some other consideration, such as congestion costs[86] or product differentiation,[87] that serves to counterbalance the scale economies.[88] Indeed, the high fixed costs associated with other content-oriented industries, such as television, have not created scale economies so large as to prevent those markets from functioning properly.[89]

Finally, we come to the prerequisite that the market is a monopoly.[90] This is a very strong assumption in the model, as can be seen if one instead explores what happens if this assumption is relaxed only slightly and the market for *A* is assumed to be a duopoly. If the market is a duopoly with two firms dividing the *A* market into two segments (a_1 and a_2), it is impossible for the producer of a_1 to use tying to exert any leverage over the market for *B*. This is because any independent producer would be able to compete for not only the *B* consumers who did not also purchase $A(b_i)$, but also B *customers* who purchased *A* from the firm that did not tie (b_2). In the numerical example given above, if we assume that the duopolists divide the market equally, an independent producers would have available to it not only the 20 consumers who

---

[85] As I noted earlier, the scale economies associated with high fixed costs decays exponentially as volume increases. Thus, even when scale economies remain unexhausted in the strict sense, as volume increases they eventually become so small and the average cost curve becomes so flat as to no longer confer any competitive advantage. At most, it would allow the dominant player to engage in a very weak form of limit pricing that was only slightly above competitive levels, in which case the consumer harm would be miniscule.

[86] See Yoo, 155 U Pa L Rev at 678-80 (cited in note **Error! Bookmark not defined.**) (showing how congestion costs can prevent markets from collapsing into natural monopolies despite unexhausted economies of scale).

[87] See Christopher S. Yoo, *Copyright and Product Differentiation*, 79 NYU L. REV. 212, 248-49 (2004) (explaining how product differentiation can prevent markets from collapsing into natural monopolies despite unexhausted economies of scale).

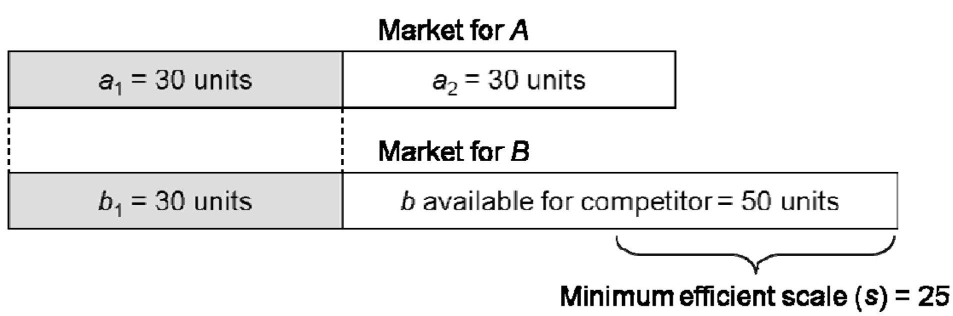
[88] There may be particular Internet services for which this may not be true. For example, auction sites that depend on bringing together large numbers of buyers and sellers may exhibit a degree of demand-side scale economies. In addition, search engines and other regimes that learn from the participation of other users may also exhibit the same quality. Whether these advantages are sufficiently large to offset any accompanying diseconomies of scale and confer a competitive advantage is an empirical question that cannot be answered *a priori*. The existence of such providers would not, however, support a blanket network neutrality rule covering all content and applications. It would instead simply support a narrow rule targeted only at content and applications for which this effect holds true.

[89] *See Implementation of Section 11(c) of the Cable Television Consumer Protection and Competition Act of 1992*, Third Report and Order, 14 FCC Rcd. 19098, 19115-16 ¶¶ 41-42 (1999) (estimating minimum efficient scale for cable television programming at 15 million homes or 18.56 percent of the market); Yoo, *supra* note 84, at 1603-07 (showing how product differentiation can offset scale economies created by high fixed costs).

[90] Whinston also analyzes a scenario in which the tying firm faces competition from an inferior provider. Whinston, *supra* note 77, at 852-54. Network neutrality proponents do not appear to rely on this particular scenario.

purchase only *B*, but also the 30 consumers who purchase both *A* and *B*, but did not purchase *A* from the tying firm. Any independent producer of *B* would have an open market of 50 units, a number well in excess of the minimum efficient scale. One could posit that the duopolists divide the market unevenly, but in that case one firm would have to control roughly 92 percent of market *A* (55/60 units) before it could exert any unilateral leverage over market *B*. Alternatively, one could instead posit that the minimum efficient scale might be 50 units. Although this would make leverage feasible, the fact that minimum efficient scale exceeded half of the market would also imply that market *B* was already a natural monopoly, in which case not only would the market not be a duopoly: tying the two products together would create efficiencies by eliminating double marginalization.[91] Of course, both duopolists could engage in tying, which would exclude independent producers. But if so, the duopolists would still engage in a degree of competition for consumers who do not also purchase $A(b_i)$. Thus, whether the duopolists would be able to profit by this depends on the conjectures about duopoly pricing.

Figure 5: Whinston's Tying Model – The Duopoly Case



The assumption that the market is a monopoly is thus a strong one that drives much of Whinston's result. If the proper geographic market is national, this assumption is clearly not

[91] See Spulber and Yoo, 107 Colum L Rev at 1838-39 & n 72 (cited in note **Error! Bookmark not defined.**) (collecting sources showing how integration between successive monopolies, successive monopoly and oligopoly, and successive oligopolies enhances consumer welfare).

met, as no provider would control more than 22 percent of the market.[92] Moreover, the assumption does not hold even if the relevant geographic market is assumed to be local. Published reports indicate that DSL is available over 80 percent of all households that can receive phone service and that cable modem service is available in over 96 percent of households that can receive cable television service.[93] Given the near ubiquity of both telephone and cable service, these numbers suggest that roughly 80 percent of the nation is served by at least two last-mile broadband Internet providers.[94] The rapid growth of wireless as a broadband platform promises to increase the competitiveness of this space still further.[95] The Whinston model by itself thus provides only weak support for a network neutrality mandate. Even under the most expansive approach to market definition, the prerequisite that the market be a monopoly limits it to at most 20 percent of the country. The other preconditions limit it still further.

The foregoing discussion highlights the narrowness of the factual preconditions that must be satisfied for Whinston's model to apply. It is for this reason that Stan Liebowitz and Stephen Margolis call Whinston's theory an example of the "Goldilocks theory of tie-in sales."[96] Simply put, everything must be just right for anticompetitive effects that Whinston identified to arise.

---

92 Yoo, *What Can Antitrust Contribute*, *supra* note 36, at 514.

93 FCC Indus. Analysis & Tech. Div., Wireline Competition Bur., High-Speed Services of Internet Access: Status as of June 30, 2007, at 3 (Mar 2008), *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-280906A1.pdf; Leichtman Research Group, Inc., Research Notes 1Q 2008, at 5, *available at* http://www.leichtmanresearch.com/research/notes03_2008.pdf (reporting that broadband Internet is available in 89 percent of homes passed by telephone wires and 99 percent of homes passed by cable television wires); Nat'l Cable & Telecomm. Ass'n, Industry Statistics, *available at* http://www.ncta.com/Statistic/Statistic/Statistics.aspx (reporting that 117.7 million of 123.4 million houses in which cable television is available can receive cable modem service or roughly 95.4 percent).

94 See Alfred E. Kahn, *Telecommunications: The Transition from Regulation to Antitrust*, 5 J. ON TELECOMM. & HIGH TECH. L. 159, 160 n 2 (2006) (performing a similar estimate based on earlier data).

95 FCC Indus. Analysis & Tech. Div., *supra* note 26, at 6 tbl. 1 (indicating that wireless has skyrocketed from having no subscribers as of the beginning on of 2005 to controlling 31 percent of the market for high-speed lines as of June 2009).

96 Stan J. Liebowitz and Stephen E. Margolis, *Bundles of Joy: The Ubiquity and Efficiency of Bundles in New Technology Markets*, 5 J. COMPETITION L & ECON. 1, 12 (2008).

As such, it is a classic example of exemplifying theory.[97] Even more importantly, Whinston acknowledges that his model does not consider whether tying might give rise to efficiencies. Thus, Whinston himself drew no conclusions about whether the type of tying he modeled would benefit or harm consumers and cautioned that his model would not support categorical governmental intervention.[98]

This is not to understate Whinston's contribution. His work represents perhaps the most important exemplifying theory used to rebut Chicago School scholars' calls for treating tying as per se legal. At the same time, it would be a mistake to attempt to use Whinston's theory as the basis for an argument in favor of a per se rule cutting in the opposite direction. As noted above, a rule declaring tying illegal per se would depend on a showing that the practice nearly always harms consumers. Whinston's analysis is an exemplifying theory that simply identifies one set of circumstances in which tying could harm consumers without making any attempt to determine the frequency with which tying would actually harm consumers.

CONCLUSION

The Internet is incredibly complex technologically, with interactions and innovations that are so myriad, multifaceted, and ever changing that no one has a complete picture of the scope of the problems. Any regulatory regime would need to be sufficiently flexible and dynamic to take all of these differences into account. At the same time, the regulatory regime must provide sufficiently clear guidance to the business community and provide a means for disposing of claims that have no chance to succeed quickly and easily.

---

97 *See* Keith N. Hylton and Michael Salinger, *Tying Law and Policy: A Decision-Theoretic Approach*, 69 ANTITRUST L.J. 469, 497 (2001) (describing Whinston's model as an exemplifying theory).

98 Whinston, *supra* note 77, at 855-56.

The approach propose above attempts to balance these concerns by requiring proof that the actor in question has both the ability and the incentive to engage in the allegedly anticompetitive conduct, that the practice actually reduces consumer welfare, and that any proposed remedies be specifically tailored to the theory of anticompetitive conduct alleged. The empirical record strongly favors placing the burden of proof on the party challenging the practice, which would provide the room for experimentation that innovation needs to succeed.